

1 Inspection Apparatus and Method

2  
3 The present invention relates to the inspection of  
4 objects including vehicles and in particular to the  
5 provision of accurate visual information from the  
6 underside of a vehicle or other object.

7  
8 Visual under vehicle inspection is of vital importance in  
9 the security sector where it is required to determine the  
10 presence of foreign objects on the underside of vehicles.  
11 Several systems currently exist which provide the means  
12 to perform such inspections.

13  
14 The simplest of these systems involves the use of a  
15 mirror placed on the end of a rod. In this case, the  
16 vehicle must be stationary as the inspector runs the  
17 mirror along the length of the car performing a manual  
18 inspection. Several problems exist with this set-up.  
19 Firstly, the vehicle must remain stationary for the  
20 duration of the inspection. The length of time taken to  
21 process a single vehicle in this way can lead to selected  
22 vehicles being inspected, as opposed to all vehicles.

1 Furthermore, it is difficult to obtain a view of the  
2 entire vehicle underside including the central section.  
3 Vitally, this could lead to an incomplete inspection and  
4 increased security risk.

5  
6 In order to combat these problems several camera based  
7 systems currently exist which either simply display the  
8 video live, or capture the vehicle underside onto  
9 recordable media for subsequent inspection. One such  
10 system involves the digging of a trench into the road. A  
11 single camera and mirror system is positioned in the  
12 trench, in such a way as to provide a complete view of  
13 the vehicle underside as it drives over. The trench is  
14 required to allow the camera and mirror system to be far  
15 enough away from the underside of the vehicle to capture  
16 the entire underside in a single image. This allows a far  
17 easier and more reliable inspection than the mirror on  
18 the rod. The main problems with this system lie with the  
19 requirement for a trench to be excavated in the road  
20 surface. This makes it expensive to install, and means  
21 that it is fixed to a specific location.

22  
23 More portable systems exist which utilise multiple  
24 cameras built into a housing similar in shape to a speed  
25 bump. These have the advantage in that they may be  
26 placed anywhere with no restructuring of the road surface  
27 required. However, these systems currently display the  
28 video footage from the multiple cameras on separate  
29 displays, one for each camera. An operator therefore has  
30 to study all the video feeds simultaneously as the car  
31 drives over the cameras. The task of locating foreign  
32 objects using this type of system is made difficult by  
33 the fact that the car is passing close to the cameras.  
34 This causes the images to change rapidly on each of the

1 camera displays, making it more likely that any foreign  
2 object would be missed by the operator.

3  
4 It is an object of the present invention to provide a  
5 system which provides an image of the entire underside of  
6 the vehicle, whilst at the same time being portable and  
7 requiring no structural alterations to the road in order  
8 to operate.

9  
10 In accordance with a first aspect of the present  
11 invention there is provided an apparatus for inspecting  
12 the under side of a vehicle, the apparatus comprising:

13 a plurality of cameras located at predetermined  
14 positions and angles relative to one another, the cameras  
15 pointing in the general direction of the area of an  
16 object to be inspected; and

17 image processing means provided with

18 (i) a first module for calibrating the cameras and  
19 for altering the perspective of image frames from  
20 said cameras and

21 (ii) a second module for constructing an accurate  
22 mosaic from said altered image frames.

23  
24 Preferably, the plurality of cameras are arranged in an  
25 array. More preferably, the array is a linear array.

26  
27 In use the apparatus of the present invention may be  
28 placed at a predetermined location facing the underside  
29 of the object to be inspected, typically a vehicle with  
30 the vehicle moving across the position of the stationary  
31 apparatus.

32  
33 Preferably the cameras have overlapping fields of view.

34

1 Preferably, the first module is provided with camera  
2 positioning means which calculate the predetermined  
3 position of each of said cameras as a function of the  
4 camera field of view, the angle of the camera to the  
5 vertical and the vertical distance between the camera and  
6 the position of the vehicle underside or object to be  
7 inspected.

8  
9 Preferably, camera perspective altering means are  
10 provided which apply an alteration to the image frame  
11 calculated using the angle information from each camera.

12  
13 Preferably, the images from each of said cameras are  
14 altered to the same scale.

15  
16 More preferably, the camera perspective altering means  
17 models a shift in the angle and position of each camera  
18 relative to the others and determines an altered view  
19 from the camera.

20  
21 The perspective shift can be used to make images from  
22 each camera appear to be taken from an angle normal to  
23 the object to be inspected or vehicle underside.

24  
25 Preferably, the camera calibration means is adapted to  
26 correct spherical lens distortion and/or non-equal  
27 scaling of pixels and/or the skew of two image axes from  
28 the perpendicular.

29  
30 Preferably, the second module is provided with means for  
31 comparing images in sequence which allows the images to  
32 be overlapped. More preferably, a Fourier analysis of  
33 the images is conducted in order to obtain the  
34 translation of x and y pixels relating the images.

1  
2 In accordance with a second aspect of the present  
3 invention there is provided a method of inspecting an  
4 area of an object, the method comprising the steps of:

- 5  
6 (a) positioning at least one camera, taking n image  
7 frames, proximate to the object  
8 (b) acquiring a first frame from the at least one  
9 camera  
10 (c) acquiring the next frame from said at least one  
11 camera  
12 (d) applying calibration and perspective alterations to  
13 said frames  
14 (e) calculating and storing mosaic parameters for said  
15 frames  
16 (f) repeat steps c to e n-1 times  
17 (g) mosaicing together the n frames from said at least  
18 one camera into a single mosaiced image.

19  
20 Preferably, the object is the underside of a vehicle.

21  
22 Preferably, a plurality of cameras is provided, each  
23 located at predetermined positions and angles relative to  
24 one another, the cameras pointing in the general  
25 direction of the object.

26  
27 Preferably, the predetermined position of each of said  
28 cameras is calculated as a function of the camera field  
29 of view and/or the angle of the camera to the vertical  
30 and/or the vertical distance between the camera and the  
31 position of the vehicle underside.

32  
33 Preferably, images from each of said cameras are altered  
34 to the same scale.

1  
2 Preferably, perspective alteration applies a correction  
3 to the image frame calculated using relative position and  
4 angle information from each camera.

5  
6 More preferably, perspective alteration models a shift in  
7 the angle and position of each camera relative to the  
8 others and determines the view therefrom.

9  
10 The perspective shift can be used to make images from  
11 each camera appear to be taken from an angle normal to  
12 the object.

13  
14 Preferably, calibration of the at least one camera  
15 corrects spherical lens distortion and/or non-equal  
16 scaling of pixels and/or the skew of two image axes from  
17 the perpendicular.

18  
19 Preferably, mosaicing the images comprises comparing  
20 images in sequence, applying fourier analysis to the said  
21 images in order to obtain the translation in x and y  
22 pixels relating the images.

23  
24 Preferably, the translation is determined by  
25 (a) Fourier transforming the original images  
26 (b) Computing the magnitude and phase of each of the  
27 images  
28 (c) Subtracting the phases of each image  
29 (d) Averaging the magnitudes of the images  
30 (e) Inverse Fourier transforming the result to produce a  
31 correlation image.  
32

1 Preferably the positioning of the at least one camera  
2 proximate to the vehicle underside is less than the  
3 vehicle's road clearance.

4  
5 Advantageously, the present invention can produce a still  
6 image rather than the video. Therefore, each point on  
7 the vehicle underside is seen in context with the rest of  
8 the vehicle. Also, any points of interest are easily  
9 examinable without recourse to the original video  
10 sequence.

11  
12 In accordance with a third aspect of the present  
13 invention there is provided a method of creating a  
14 reference map of an object, the method comprising the  
15 steps of obtaining a single mosaiced image, selecting an  
16 area of the single mosaiced image and recreating or  
17 selecting the frame from which said area of the mosaiced  
18 image was created.

19  
20 Preferably, the area of the single mosaiced image is  
21 selected graphically by using a cursor on a computer  
22 screen.

23  
24 The present invention will now be described by way of  
25 example only with reference to the accompanying drawings  
26 of which:

27       FIGURE 1 is a schematic diagram for the high level  
28       processes of this invention;

29  
30       FIGURE 2 shows the camera layouts for one half of  
31       the symmetrical unit in the preferred embodiment;  
32

FIGURE 3 is schematic of the camera pose alteration required to correct for perspective in each of the image frames by;

FIGURE 4 demonstrates the increase in viewable achieved when the camera is angled; and

FIGURE 5 is a flow diagram of the method applied when correcting images for the sensor roll and pitch data concurrently with the camera calibration correction.

A mosaic is a composite image produced by stitching together frames such that similar regions overlap. The output gives a representation of the scene as a whole, rather than a sequential view of parts of that scene, as in the case of a video survey of a scene. In this case, it is required to produce a view of acceptable resolution at all points of the entire underside of a vehicle in a single pass. In this example of the present invention, this is accomplished by using a plurality of cameras arranged in such a way as to achieve full coverage when the distance between the cameras and vehicle is less than the vehicle's road clearance.

An example of such a set up using five cameras is provided in figure 2; the width of the system being limited by the wheel base of the vehicle. This diagram shows one half of the symmetric camera setup with the centre camera, angled  $0^\circ$  to the vertical, to the right of the figure.

The notation used in figure 1 is defined as follows:

$L_u$  = Width of unit.



1  $L_e$  = Maximum expected width of vehicle.

2  $h$  = Minimum expected height from the camera lenses  
3 to the vehicle.

4  $\tau$  = True field of view of camera.

5  $\tau'$  = Assumed field of view of camera, where  $\tau' = \tau - \delta\tau$  and  
6  $0 < \delta\tau < \tau$ .

7  $\theta_i$  = Angles of outer cameras to the vertical, where  
8  $i=1,2$ .

9  $L_i$  = Distances of outer cameras from the central  
10 camera, where  $L_1 < L_2 < L_u/2$ .

11  
12 In this notation an assumed field of view  $\tau'$  is used, as  
13 opposed to the true field of view  $\tau$ , the reason for this is  
14 twofold. Firstly, it provides a redundancy in the cross-  
15 camera overlap regions ensuring the vehicle underside is  
16 captured in its entirety. Secondly, in the case of a  
17 vehicle that is of maximal width, the use of  $\tau$  in the  
18 positioning calculations will lead to resolution problems  
19 at the outer edge of the vehicle. These problems become  
20 evident when the necessary image corrections are  
21 performed.

22  
23 Knowing  $L_e$ ,  $h$ ,  $\tau'$ , and  $L_2$ , then  $\theta_2$  may be calculated as

$$\theta_2 = \tan^{-1} \left[ \frac{L_e - 2L_2}{2h} \right] - \frac{\tau'}{2}$$

26  
27 Using this geometry  $\theta_1$  cannot be determined analytically.  
28 It is therefore calculated as the root of the following  
29 equation through use of a root finding technique such as  
30 the bisection method

31

$$\tan\left(\frac{\tau'}{2} + \theta_1\right) + \tan\left(\frac{\tau'}{2} - \theta_1\right) + \left[ \tan\left(\frac{\tau'}{2}\right) + \tan\left(\frac{\tau'}{2} - \theta_2\right) - \frac{L_2}{h} \right] = 0$$

2

3 Following this the distance  $L_1$  is calculated as

4

$$L_1 = h \left[ \tan\left(\frac{\tau'}{2}\right) + \tan\left(\frac{\tau'}{2} - \theta_1\right) \right]$$

6

7 The use of these equations ensures the total coverage of  
8 the underside of a vehicle whose dimensions are within  
9 the given specifications.

10

11 In estimating the interframe mosaicing parameters of  
12 video sequences there are currently two types of method  
13 available. The first uses feature matching within the  
14 image to locate objects and then to align the two frames  
15 based on the positions of common objects. The second  
16 method is frequency based, and uses the properties of the  
17 Fourier transform.

18

19 Given the volume of data involved (a typical capture rate  
20 being 30 frames per second) it is important that a  
21 technique which will provide us with a fast data  
22 throughput is utilised, whilst also being highly accurate  
23 in a multitude of working environments. In order to  
24 achieve these goals, the correlation technique based on  
25 the frequency content of the images being compared is  
26 used. This approach has two main advantages:

27

28 1. Firstly, regions that would appear relatively  
29 featureless, that is those not containing strong  
30 corners, linear features, and such like, still contain  
31 a wealth of frequency information representative of the  
32 scene. This is extremely important when mosaicing

1 regions of the seabed for example, as definite features  
2 (such as corners or edges) may be sparsely distributed;  
3 if indeed they exist at all.

4 2. Secondly, the fact that this technique is based on the  
5 Fourier transform means that it opens itself  
6 immediately to fast implementation through highly  
7 optimized software and hardware solutions.

8  
9 Implementation steps in order of their application will  
10 now be discussed.

11  
12 All cameras suffer from various forms of distortion.  
13 This distortion arises from certain artefacts inherent to  
14 the internal camera geometric and optical characteristics  
15 (otherwise known as the intrinsic parameters). These  
16 artefacts include spherical lens distortion about the  
17 principal point of the system, non-equal scaling of  
18 pixels in the x and y-axis, and a skew of the two image  
19 axes from the perpendicular. For high accuracy mosaicing  
20 the parameters leading to these distortions must be  
21 estimated and compensated for. In order to correctly  
22 estimate these parameters images taken from multiple  
23 viewpoints of a regular grid, or chessboard type pattern  
24 are used. The corner positions are located in each image  
25 using a corner detection algorithm. The resulting points  
26 are then used as input to a camera calibration algorithm  
27 well documented in the literature.

28  
29 The estimated intrinsic parameter matrix  $A$  is of the form

$$A = \begin{bmatrix} \alpha & \gamma & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix}$$

1 where  $\alpha$  and  $\beta$  are the focal lengths in x and y pixels  
2 respectively,  $\gamma$  is a factor accounting for skew due to  
3 non-rectangular pixels, and  $(u_0, v_0)$  is the principle point  
4 (that is the perpendicular projection of the camera focal  
5 point onto the image plane).

6  
7 A prerequisite for using the Fourier correlation  
8 technique is that consecutive images must match under a  
9 strictly linear transformation; translation in x and y,  
10 rotation, and scaling. Therefore the assumption is made  
11 that the camera is travelling in a direction normal to  
12 that in which it is viewing. In the case of producing an  
13 image of the underside of a vehicle, this assumption  
14 means that the camera is pointing strictly upward at all  
15 times. The fact that this may not be the case with the  
16 outer cameras leads to the perspective corrected images  
17 being used in the processing.

18  
19 This is accomplished by modelling a shift in the camera  
20 pose and determining the normal view from the captured  
21 view. In order to accomplish this, the effective focal  
22 distance of the camera is required. This value is needed  
23 in order to perform for the projective transformation  
24 from 3D coordinates into image pixel coordinates, and is  
25 gained during the intrinsic camera parameter estimation.  
26 Figure 3 shows a diagram of this pose shift.

27  
28 When correcting for perspective, the new camera position  
29 is at the same height as the original viewpoint, not the  
30 slant range distance. Thus all of the images from each  
31 of the cameras are corrected to the same scale.

32  
33 For each image comparison of images from the chosen  
34 camera, it is assumed that there is no rotation or

1 zooming differences between the frames. This way only  
2 the translation in x and y pixels need be estimated.  
3 Having obtained the necessary parameters of the  
4 differences in position of the two images, they can be  
5 placed in their correct relative positions. The next  
6 frame is then analysed in a similar manner and added to  
7 the evolving mosaic image. A description of the  
8 implementation procedures used in this invention for  
9 translation estimation in Fourier space will now be  
10 given.

11  
12 In Fourier space, translation is a phase shift. The  
13 differences in the phase to determine the translational  
14 shift. Let the two images be described by  $f_1(x,y)$  and  
15  $f_2(x,y)$  where  $(x,y)$  represents a pixel at this position  
16 should be utilised. Then for a translation  $(dx,dy)$  the two  
17 frames are related by

$$f_2(x,y) = f_1(x+dx, y+dy)$$

18  
19  
20  
21 The Fourier transform magnitudes of these two images are  
22 the same since the translation only affects the phases.  
23 Let our original images be of size  $(cols,rows)$ , then each of  
24 these axes represents a range of  $2\pi$  radians. So a shift  
25 of  $dx$  pixels corresponds to  $2\pi dx/cols$  shift in phase for  
26 the column axis. Similarly, a shift of  $dy$  pixels  
27 corresponds to  $2\pi dy/rows$  shift in phase for the row axis.

28  
29 To determine a translation, a Fourier transform of the  
30 original images, compute the magnitude ( $M$ ) and phases  
31 ( $\phi$ ) of each of the pixels and subtract the phases of each  
32 pixel to get  $d\phi$ . The average of the magnitudes (they

1 should be the same) and the phase differences are taken  
2 and a new set of real ( $\Re$ ) and imaginary ( $\Im$ ) values as  
3  $\Re = M \cos(d\phi)$  and  $\Im = M \sin(d\phi)$  is computed. These ( $\Re, \Im$ ) values  
4 are then inverse Fourier transformed to produce an image.  
5 Ideally, this image will have a single bright pixel at a  
6 position( $x, y$ ), which represents the translation between  
7 the original two images, whereupon a subpixel translation  
8 estimation may be made.

9  
10 An important point to consider is which camera to use in  
11 calculating the mosaicing parameters. When asking this  
12 question the primary consideration is that of overlap,  
13 and how to get the maximum effective overlap between  
14 frames. It is here that an added benefit is found to  
15 having the outer cameras angled. If the centre camera is  
16 used then the distance subtended by the view of a single  
17 frame along the central axis of that frame is

$$d_c = 2h \tan(\tau'/2)$$

18  
19  
20  
21 When the camera is rolled to an angle of  $\theta_1$  degrees to the  
22 vertical as shown in figure 2, then the distance  
23 subtended by the view of a single frame along the central  
24 axis is

$$d_1 = 2h \tan(\tau'/2) / \cos(\theta_1)$$

25  
26  
27  
28 which is greater than  $d_c$  for all  $\theta_1 \neq 0$ . This property is  
29 illustrated in figure 4.

30  
31 Care must be exercised here however as according to this  
32 argument one of the cameras at the greatest angle  $\theta_2$   
33 should be used. Two reasons count against this choice.

1 Firstly, the pixel resolution at the outer limits of the  
2 corrected image is the poorest of all the imaged areas.  
3 Secondly, and most importantly, due to the enforced  
4 redundancy in the coverage, and that most vehicles will  
5 fall short of the maximum width limits, the outer region  
6 of this image (that which should correspond to the  
7 maximum overlap) does not view the underside of the  
8 vehicle at all. In this case most of the image will  
9 contain stationary information. For these reasons it is  
10 recommended that one of the cameras angled at  $\theta_1$  degrees  
11 should be used.

12  
13 Given the mosaicing parameters, the final stage of the  
14 process is to stitch the corrected images into a single  
15 view of the underside of the vehicle. The first point to  
16 stress here is that mosaicing parameters are only  
17 calculated along the length of the vehicle, not between  
18 each of the cameras. The reason for this is that there  
19 will be minimal, as well as variable, overlap between  
20 camera views. These problems mean that any mosaicing  
21 attempted between the cameras will be unreliable at best.  
22 For this reason each of the camera images at a given  
23 instant in time are cropped to an equal number of rows,  
24 and subsequently placed together in a manner which  
25 assumes no overlap.

26  
27 These image strips are then stitched together along the  
28 length of the car using the calculated mosaicing  
29 parameters, providing a complete view of the underside of  
30 the vehicle in a single image. This stitching is  
31 performed in such a way that the edges between strips are  
32 blended together. In this blending the higher resolution  
33 central portions of each frame are given a greater  
34 weighting.

1  
 2 A final point to note here is that when the final  
 3 stitched result is calculated, each of the pixel values  
 4 is interpolated directly from the captured images. This  
 5 is achieved through use of pixel maps relating the pixel  
 6 positions in the corrected image strips directly to the  
 7 corresponding sub-pixel positions in the captured images.  
 8 The advantage of adopting this approach is that only a  
 9 single interpolation stage is used. This has the effect  
 10 of not only reducing memory requirements and saving  
 11 greatly on processing time, but also the resultant image  
 12 is of a higher quality than if multiple interpolation  
 13 stages had been used; a schematic for this process is  
 14 provided in figure 5. The process of generating the  
 15 pixel maps correcting for camera calibration and  
 16 perspective correction are combined mathematically in the  
 17 following way.

18  
 19 If  $\underline{u}$  is the corrected pixel position, the corresponding  
 20 position in the reference frame of the camera, normalised  
 21 according the camera focal length in y pixels ( $\beta$ ) and  
 22 centred on the principle point  $(u_0, v_0)$ , is  
 23  $\underline{c}' = [(c_1'', c_2'', c_3'')/c_4'' - (u_0, v_0)]/\beta$  where  $\underline{c}'' = PR_y R_x P^{-1} \underline{u}$ . The pitch  
 24 and roll are represented by the rotation matrices  $R_x$  and  
 25  $R_y$ , respectively, with P being the perspective projection  
 26 matrix which maps real world coordinates onto image  
 27 coordinates. Following this the pixel position in the  
 28 captured image is calculated as  $\underline{c} = A \tau_c \underline{c}'$ . The scalar  $\tau_c$   
 29 represents the radial distortion applied at the camera  
 30 reference frame coordinate  $\underline{c}'$ . The matrix A is as  
 31 defined previously.

32



1 The apparatus and method of the present invention may  
2 also be used to re-create each of the images from which  
3 the mosaiced image was created.

4  
5 Once the mosaiced image has been created, it can be  
6 displayed on a computer screen. If an area of the image  
7 is selected on the computer screen using the computer  
8 cursor, the method and apparatus of the present invention  
9 can determine the image from which this part of the  
10 mosaic was created and can select this image frame for  
11 display on the screen. This can be achieved by  
12 identifying and selecting the correct image for display  
13 or by reversing the mosaicing process to return to the  
14 original image.

15  
16 In practice, this feature may be used where a particular  
17 part of an object is of interest. If for example, the  
18 viewer wishes to inspect a part of the exhaust on the  
19 underside of a vehicle then the image containing this  
20 part of the exhaust can be recreated.

21  
22 Improvements and modifications may be incorporated herein  
23 without deviating from the scope of the invention.